Los Alamos NATIONAL LABORATORY EST.1943

Machine Learning for the Grid

D. Deka, S. Backhaus & *M. Chertkov* +

A. Lokhov, S. Misra, M. Vuffray and K. Dvijotham

DOE/OE & LANL (Grid Science) + GMLC (1.4.9 + 2.0)

D.Deka

S. Backhaus

M. Vuffray

A. Lokhov

S. Misra

K. Dvijotham (Caltech)

M. Chertkov

- Intro: Overview of Challenges and Approaches
- Technical Intro: Direct and Inverse Stochastic Problem
    - –Machine Learning for Grid Operations
- Machine Learning for Distribution Grid
- Machine Learning for Transmission Grid
- Graphical Models & New Physics=Grid Informed Learning Tools

Data Analytics can improve resiliency in the Dynamic Grid

Changes in the modern Grid:

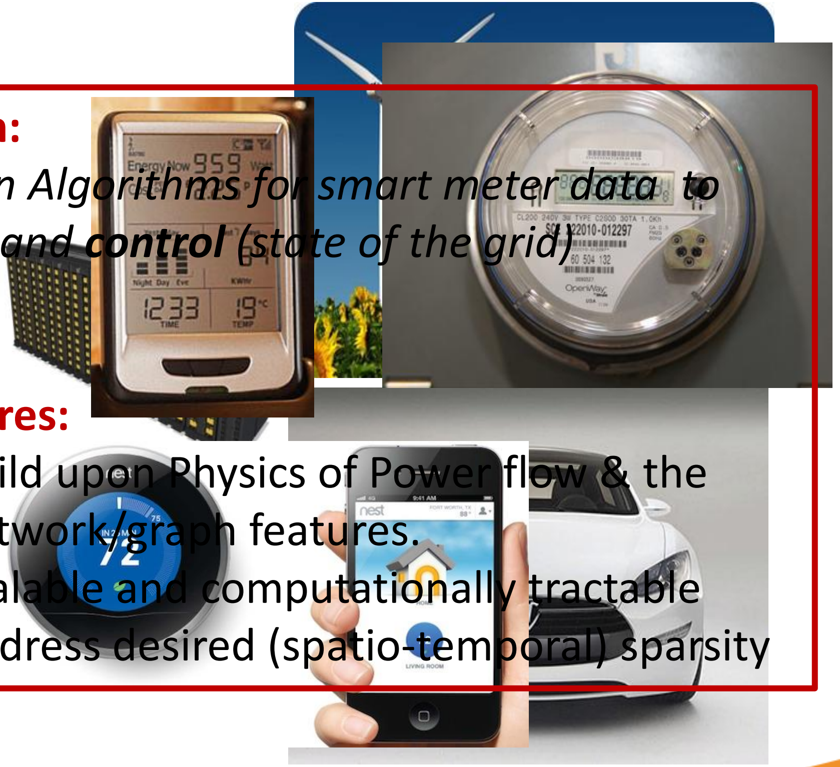
- Penetration of Renewables
- Storage devices
- Loads becomes active (not controlled)

Challenges

- Strong fluctuations/uncertainty
- Needs real-time observability, control
- Millions of devices, many entities

New (available) Solutions

- Hardware:
  Smart meters, PMUs, micro-PMUs
- Software/New algorithms:
  Machine Learning, IoT



Vision:

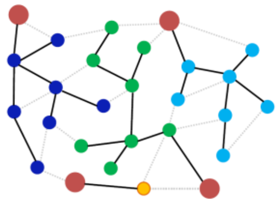
Design Algorithms for smart meter data ***to learn and control*** *(state of the grid)*

Features:

- Build upon Physics of Power flow & the network/graph features.
- Scalable and computationally tractable
- Address desired (spatio-temporal) sparsity

- Intro: Overview of Challenges and Approaches
- Technical Intro: Direct and Inverse Stochastic Problem
  –Machine Learning for Grid Operations
- Machine Learning for Distribution Grid
- Machine Learning for Transmission Grid
- Graphical Models & New Physics=Grid Informed Learning Tools

Grid should operate in spite of uncertainty & fluctuations



uncertainty:

- Graph Layout (switching of lines) + other +/- variables (transformers)
- State Estimation (consumption & production)
    - Deterministic static & dynamic models (e.g. relating s=(p,q) to v)
    - Probabilistic (statistical) models =>

$$s_j = \sum_{k\sim j} v_j \left(\frac{v_j - v_k}{z_{jk}}\right)^*$$

Power Flow Eqs.

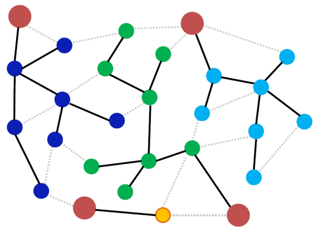
fluctuations:

- Renewable generators (wind & solar)
- loads (especially if active = involved in Demand Response)

Direct Deterministic Problem: Power Flow (static/minutes)

Given:

- operational grid=graph, inductances/resistances
- injections/consumptions (for example)



Compute:

- power flows over lines
- voltages
- phases

$$s_j = \sum_{k\sim j} v_j \left(\frac{v_j - v_k}{z_{jk}}\right)^*$$

Power Flow Eqs.

Direct Stochastic Problem: Power Flow (static/minutes)

Given:

- operational grid=graph, inductances/resistances
- Probability distribution (statistics) of injections/consumptions (for example)

  -- samples are assumed drawn (from the probability distribution), e.g. i.i.d.

Compute statistics of:

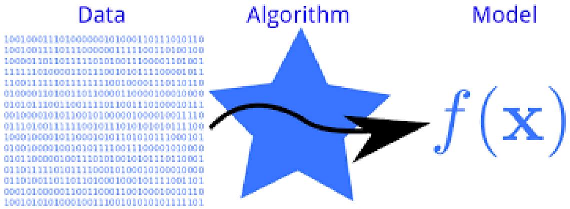
- power flows over lines
- voltages
- phases

$$s_j = \sum_{k\sim j} v_j \left(\frac{v_j - v_k}{z_{jk}}\right)^*$$

Power Flow Eqs.

joint & marginal probability distributions

Inverse Stochastic Problem: Power Flow (static/minutes)

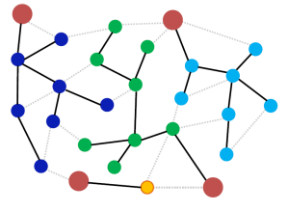


$$s_j = \sum_{k\sim j} v_j \left(\frac{v_j - v_k}{z_{jk}}\right)^*$$

Power Flow Eqs.

Given:

- ~~operational~~ grid=graph, inductances/resistances
- snapshots/measurements of power flows, voltages, phases
- parametrized representation for statistics of injections/consumptions, e.g. Gaussian & white



Infer/*Learn*:

- parameters for statistics of the injection/consumption
- operational grid=graph

Sample/Predict:

- configurations of injection/consumption
  => direct problem (compute)

Machine Learning for the Grid (at least some part) =
Automatic Solution of the Inverse Grid Problem(s)

Many flavors:

- static vs dynamic
- transmission vs distribution
- blind (black box) vs grid/physics informed
- samples vs moments (sufficiency)
- principal limits (IT) vs efficient algorithms
- ML for model reduction
- individual devices vs ensemble learning

[focus only on some of these ``complexities” in the talk]

- Intro: Overview of Challenges and Approaches
- Technical Intro: Direct and Inverse Stochastic Problem
  - –Machine Learning for Grid Operations
- **Machine Learning for Distribution Grid**
- Machine Learning for Transmission Grid
- Graphical Models & New Physics=Grid Informed Learning Tools

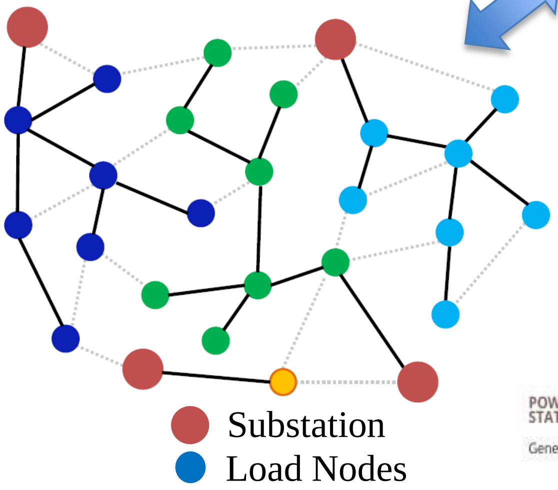
Machine Learning for Distribution Grid

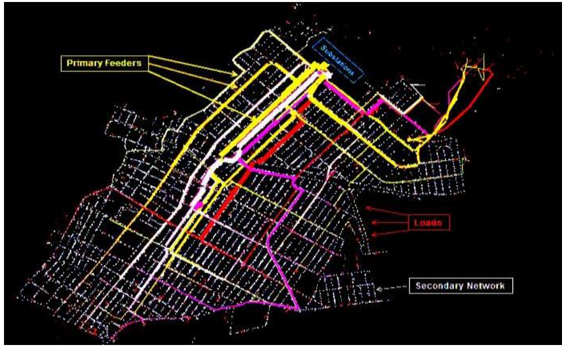
D. Deka, S. Backhaus, MC
arxiv:1502.07820, 1501.04131, +

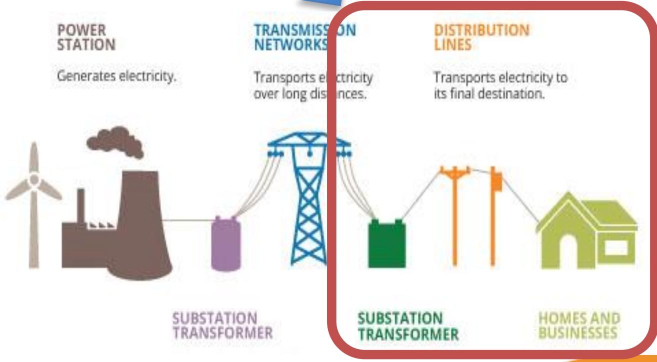
Learn
- Switch statuses
- Load statistics, line impedances

Challenges
- Nodal Measurements (voltages)
- Missing Nodes
- Information limited to households

Key Ideas
- Operated Radial structure
- Linear-Coupled power flow model
- Graph Learning tricks







Machine Learning for Distribution Grid

D. Deka, S. Backhaus, MC
arxiv:1502.07820, 1501.04131, +

Linear-Coupled power flow model:

$$P_a + \hat{i}Q_a = \sum_{(a,b)\text{ is edge}} V_a e^{\hat{i}\theta_a}(V_a e^{-\hat{i}\theta_a} - V_b e^{-\hat{i}\theta_b})/(R_{ab} - \hat{i}X_{ab})$$

$$V_a \approx 1,\ \ \theta_a - \theta_b \approx 0$$

equivalent to LinDistFlow (Baran-Wu)

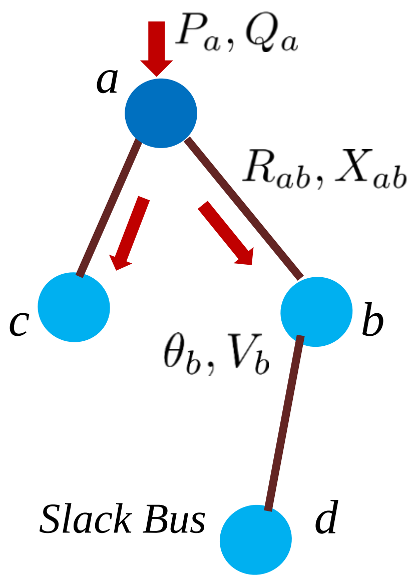
$$\boxed{\theta = H^{-1}_{1/X}P - H^{-1}_{1/R}Q, \quad V = H^{-1}_{1/R}P + H^{-1}_{1/X}Q}$$

$$H_{1/R} = M^T R^{-1} M$$

reduced Laplacian matrix

reduced Incidence matrix

Inverse Matrices are computable explicitly on trees



Machine Learning for Distribution Grid

D. Deka, S. Backhaus, MC
arxiv:1502.07820, 1501.04131, +

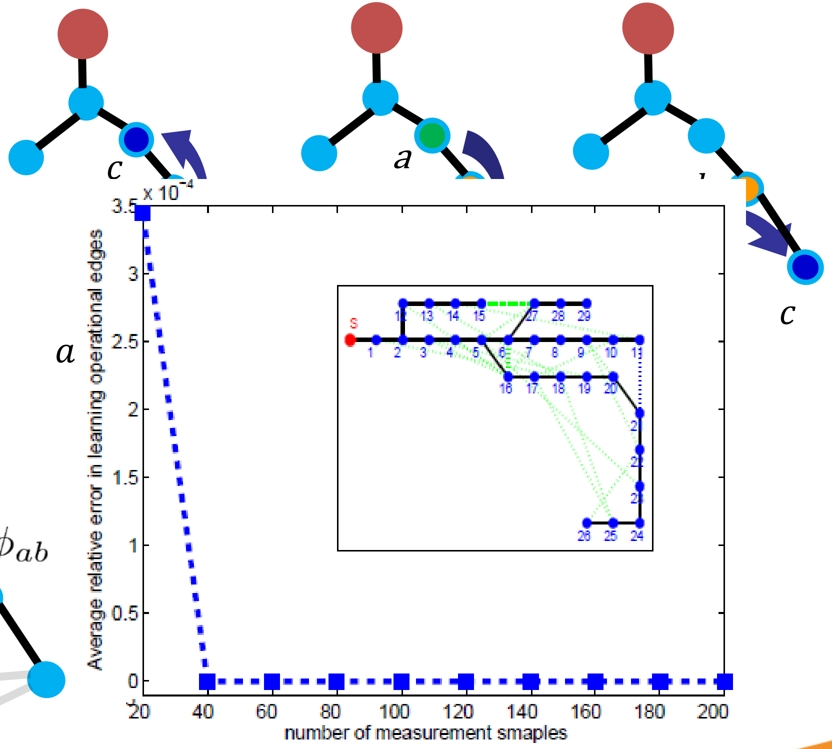
Key Idea:

- Use *variance of voltage diff.* as edge weights

$$\boxed{\phi_{ab} = \mathbb{E}[(V_a - \mu_{V_a}) - (V_b - \mu_{V_b})]^2}$$

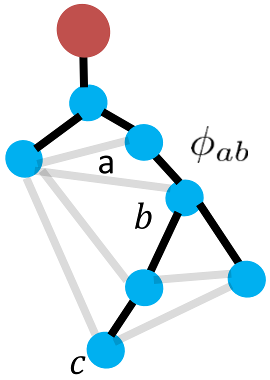
- Minimal value outputs the *nearest neighbor*

$$\phi_{ab} < \phi_{ac}$$



Learning Algorithm:

- *Min spanning tree* with variance of voltage diff. as edge weights



✓ No other information needed
✓ Low Complexity: $O(E \log E)$
✓ Can learn covariance of fluctuating loads

Machine Learning for Distribution Grid

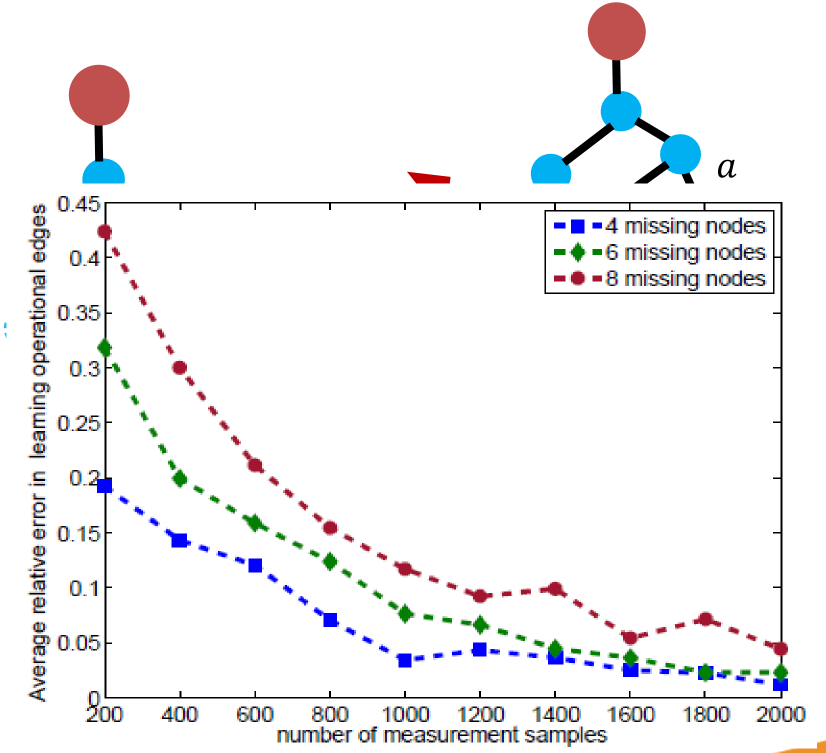
D. Deka, S. Backhaus, MC
arxiv:1502.07820, 1501.04131, +

Learning with missing nodes:

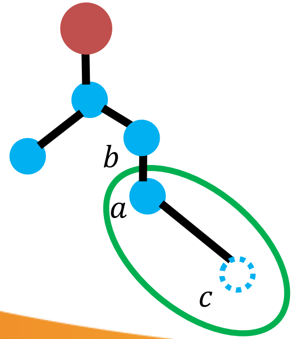
- Missing nodes separated by *2 or more hops*

Learning Algorithm:

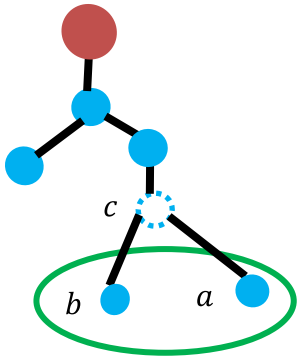
- *Min spanning tree* with available nodes
- Starting from leaf, check *missing node*

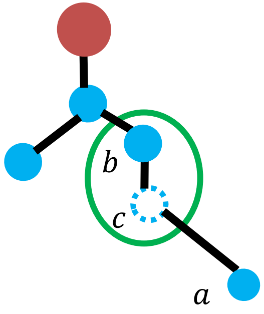


Leaf



Intermediate





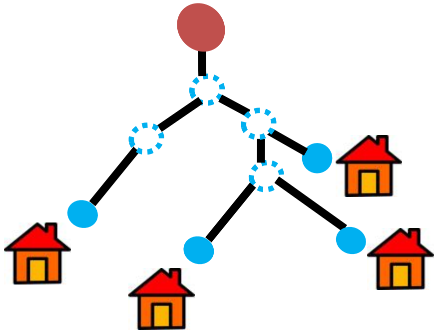
UNCLASSIFIED

Machine Learning for Distribution Grid

D. Deka, S. Backhaus, MC
arxiv:1502.07820, 1501.04131, +

Learning with missing nodes & reduced information:

- Missing nodes separated by *2 or more hops*
- Model reduction, ensemble (sampling distributions)



Extensions:

- ✓ Learn using **end-node (household)** data accounting for
  - ✓ mix of active (with control) & passive
  - ✓ dynamics of loads/motors and inverters
  - ✓ emergencies, e.g. FIDVR
- ✓ Learn 3 phase unbalanced networks
- ✓ Learn loopy grid graph
  - ✓ cities (Manhattan)
  - ✓ rich exogenous correlations (loops representing non-grid knowledge)

- ✓ Coupling to other physical infrastructures
  - gas/water distribution
  - thermal heating

  e.g. extending the learning methodology to the more general ``physical flow" networks

Recently Awarded GMLC:

Topic 1.4.9 Integrated Multi Scale Data Analytics and Machine Learning for the Grid

PIs:

Emma Stewart (LBNL)

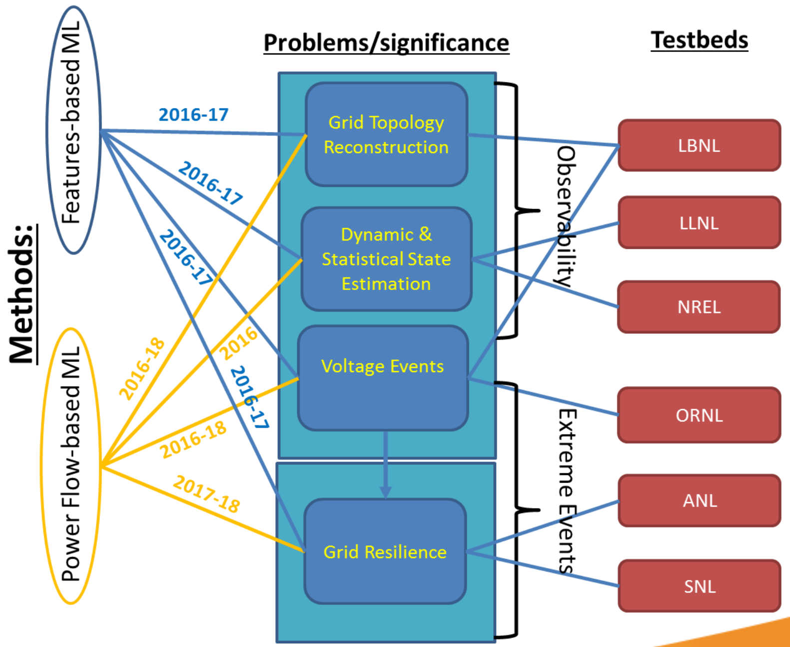
Michael Chertkov (LANL)

NL involved:

LBNL,LANL, SNL, ORNL, LLNL, NREL, ANL

- Platform
  - review
  - development,
  - data collection
- ML and Data Analytics for Visibility
- ML and Data Analytics for Resilience

Road Map of 1.4.9 (ML for distribution grids)



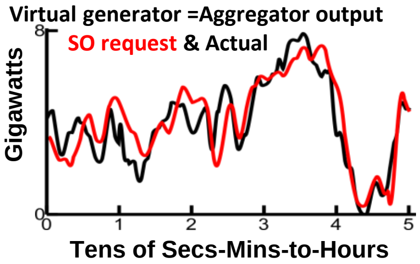
Integrating Distrib.-Level (stochastic) Loads in Frequency Control

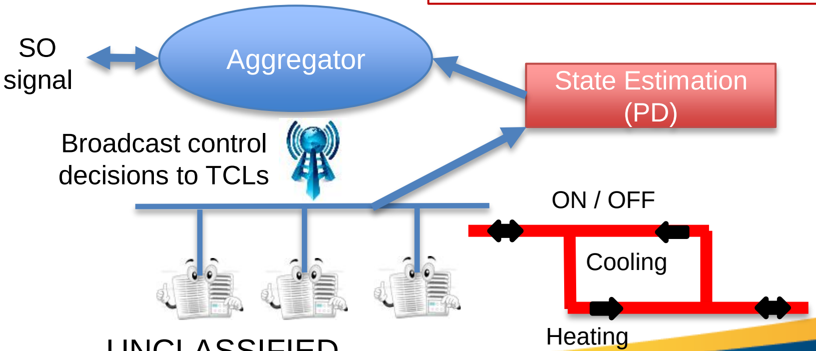
Idea: Use distribution level Demand Response (DR), specifically ensemble of Thermostatically Control Loads (TCL), to balance SO signal through Aggregator (A)

- Thousands of TCLs are aggregated
- SO->Aggregator (A)->TCLs [top-> bottom]
- Aggregator is seen (from above) as a "virtual GEN"
  Goal of the study to answer the principal question:
  - **Can A follow the SO's real-time signal as an actual GEN?**
- … and do it under "social welfare" conditions [our novel approach]:
  TCLs are controlled by the aggregator in a least intrusive way
  - broadcast of a few control signals
    (switching [stochastic] rates, temperature band)
  - probability distribution (PD) over states (temperature, +/-)
    is the control variable

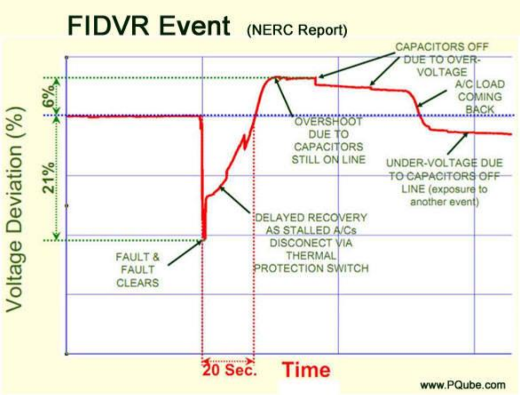
Results & Work in Progress:

- Builds on theory & simulation experience from Nonequilibrium StatMech & Control
- Stochastic/PDE/spectral methods for analysis of the PD ("driven" Fokker-Planck) were developed and cross-validated
- Ensemble Control Scheme ("second quantization"= Bellman-Hamilton-Jacobi approach for PD) is formulated … testing.



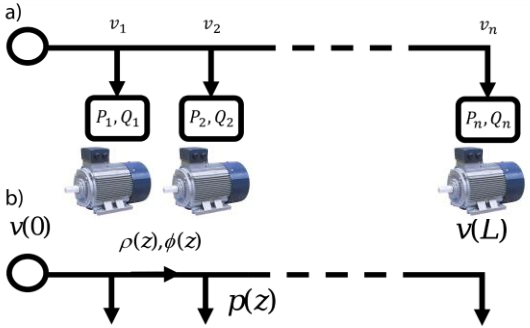


Fault-Induced Delayed Voltage Recovery

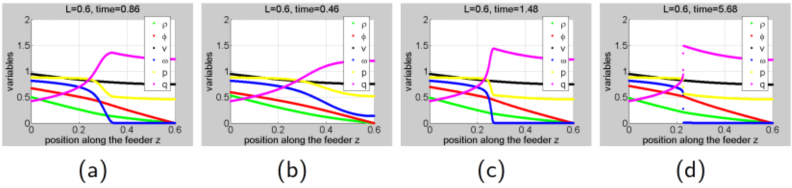


Challenges:

- Describe FIDVR quantitatively
- Learn to detect it fast
- Predict if a developing event will or will not lead to recovery? Cascade?
- Develop minimal preventive emergency controls

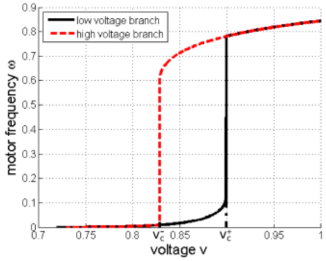


Example of a Small Fault → feeder is partially stalled (Movie Small Fault)



Results:

- Reduced PDE model was developed
- Distributed Hysteretic behavior was described
- Effects of disorder and stochasticity were analyzed
- Effect of cascading from one feeder to another and possibly further to transm. was investigated

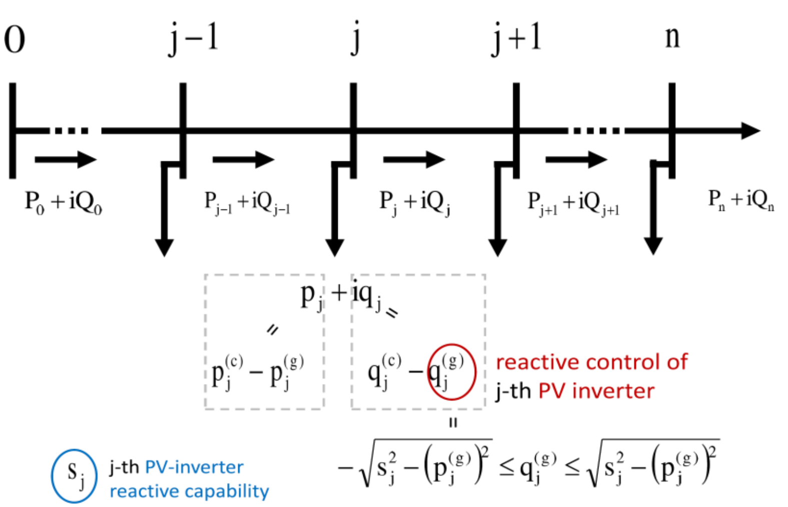


Hysteretic behavior/stalling

Work in progress:

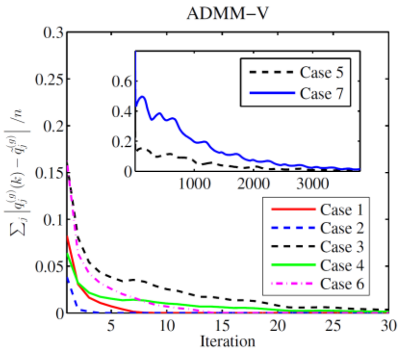
- Effects of other devices (controlled or not)
- Preventive/emergency control

Optimal Distributed Control of Reactive Power via ADMM



Challenges:

- Develop algorithm to control voltage and losses in distribution
- Do it using/exploring new degree of freedom = reactive capabilities of inverters



performance

$$\min_{q^{(g)},P,Q,V} \sum_{j=0}^{n-1} r_j \frac{P_j^2 + Q_j^2}{V_j^2} \quad = \text{Losses}$$

s.t. Power Flow Equations

$$\forall j = 1, \ldots, n:$$

$$(1-\epsilon)^2 V_0^2 \le V_j^2 \le (1+\epsilon)^2 V_0^2$$

$$\left|q_j^{(g)}\right| \le \sqrt{s_j^2 - \left(p_j^{(g)}\right)^2}.$$

Case	Loss^g	Loss^l
1	0.834	0.845
2	0.941	0.949
3	0.847	0.890
4	0.954	0.962
6	0.700	0.771

local vs global

Validated on realistic distribution circuits

Case	Nodes	PV-pen	$p^{(c)}_{max}$	$p^{(g)}$	s_{max}
1	100	100%	4 kW	1 kW	1.1 kW
2	100	50%	4 kW	1 kW	1.1 kW
3	250	50%	2.5 kW	1 kW	2.2 kW
4	250	50%	1 kW	2 kW	2.2 kW
5	200	100%	3.75 kW	0 kW	2.2 kW
6	150	85%	4 kW	0.9 kW	1.1 kW
7	150	70%	2 kW	6.5 kW	10 kW

Results:

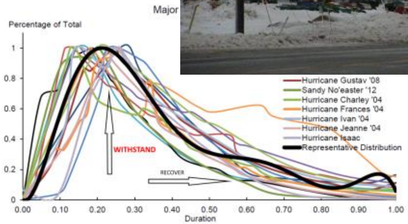
- The developed control (based on the LinDistFlow representation of the Power Flows in distribution is
  - Distributed (local measurements + communications with neighbors)
  - Efficient = implemented via powerful ADMM) (Alternative Direction Method of Multipliers

Resilient Distribution Systems (Bent, Backhaus, Yamangil, Nagarajan)

Goal: Withstand the initial impact of large-scale disruptions

Develop tools, methodologies, and algorithms to enable the design of resilient distribution systems, using

- Asset hardening
- System expansion by adding new:
  - Lines/circuit segments
  - Switching
  - Microgrid facilities
  - Microgrid generation capacity
- Binary decisions, mixed-integer programming problem



Source: Department of Energy, Office of Electrcity Delivery and Energy Reliability

Model

Given a graph $G = (V, E)$ where V and E corresponds to **node based and edge based upgrades**, respectively, and S a **set of disaster scenarios**, we want to find:

$$
\begin{array}{lll}
\min & Budget(G') & \\
\text{s.t.} & G' \subseteq G & \\
 & T_s \subseteq G' & \forall s \in S \\
 & T_s \in OperatingConditions(G) & \forall s \in S \\
 & CriticalDemand(T_s) \geq MinCriticalDemand & \forall s \in S \\
 & TotalDemand(T_s) \geq MinTotalDemand & \forall s \in S
\end{array}
$$

A simplified model

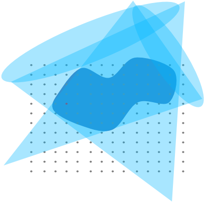
1st stage: construction variables

2nd stage: assets in use

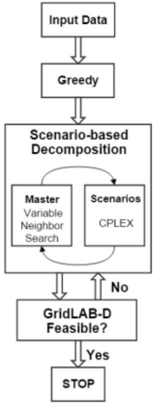
Block diagonal structure: coupling variables.

LinDistFlow: 3-phase unbalanced AC power flow

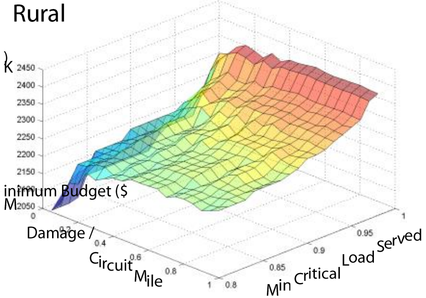
Relaxations

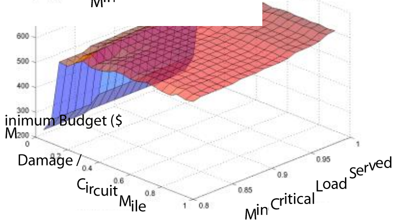


Algorithm



Rural





Observations

- Rural networks require larger resilience budgets/MW served.
  - Microgrids favored over hardened lines
- Urban budget is insensitive to critical load requirements
  - Minimal hardening of lines achieves resilience goals

UNCLASSIFIED

- Intro: Overview of Challenges and Approaches
- Technical Intro: Direct and Inverse Stochastic Problem
  - –Machine Learning for Grid Operations
- Machine Learning for Distribution Grid
- **Machine Learning for Transmission Grid**
- Graphical Models & New Physics=Grid Informed Learning Tools

Machine Learning for the Transmission Grid:

Ambient Stage

D. Deka, S. Backhaus, MC + work in progress

Learn

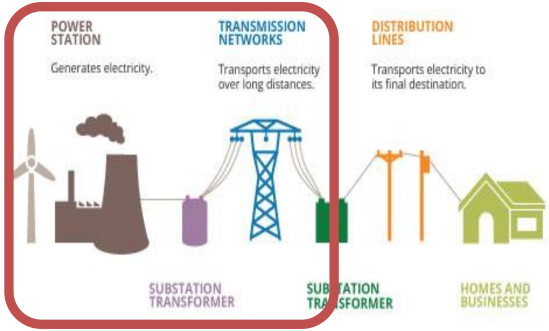
- Inertia, damping for generators
- Key parameters for (aggregated) loads (state estimation)
- Statistics of spatio-temporal fluctuations (statistical state estimation)
- Critical wave-modes (speed of propagation, damping)

Key Ideas

- Temporal scale separation:
  slow (tens of mins) vs fast (tens of secs)
- Learning stochastic ODEs – generalized stochastic swing equations
- Spatial aggregation – incorporating PDEs
- Green function approach (extending Backhaus & Liu 2011 beyond detailed balance)

Challenges

- Limited measurements
- Incorporating PMU with SCADA
- On-line requirements,
  e.g. need linear scaling algorithms



Machine Learning for the Transmission Grid:

Detection & Mitigation of Frequency Events

D. Deka, S. Backhaus, MC + work in progress

Learn

- Detect, localize & size frequency events in almost real time, utilizing ambient state estimation

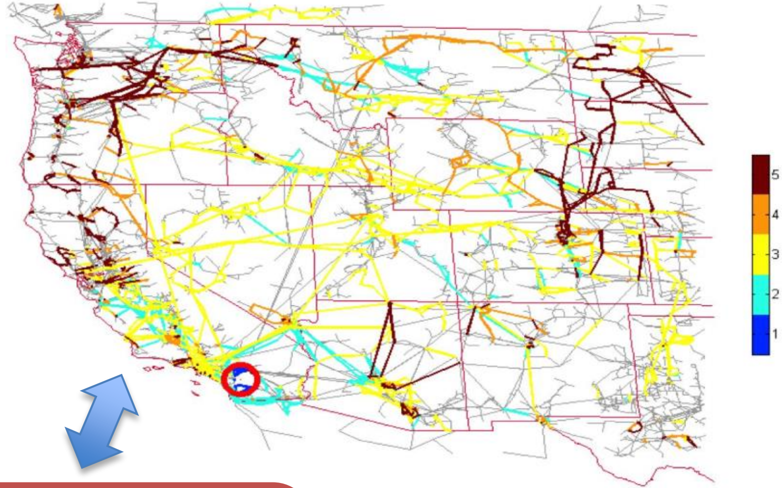
Challenges

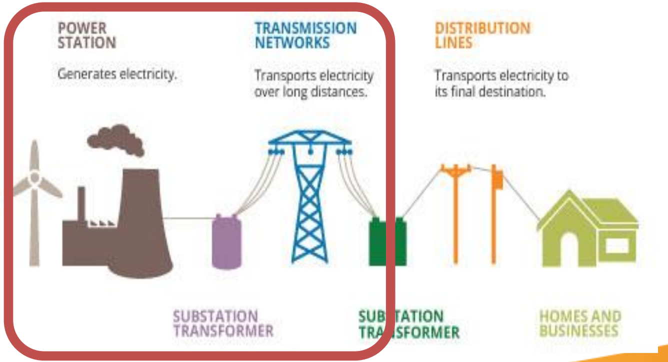
- Spatio-temporaly optimal, fast measurements
- Have a fast predictive power – is an extra control needed? when? where?

Key Ideas

- Modeling: electro-mechanical waves over 1d+ and/or 2d aggregated media, forerunner (shortest path), interference pattern

$$M\partial_t^2\theta + \tau\partial_t\theta = D\partial_x^2\theta + A\delta(t-t_0)\delta(x-x_0)$$





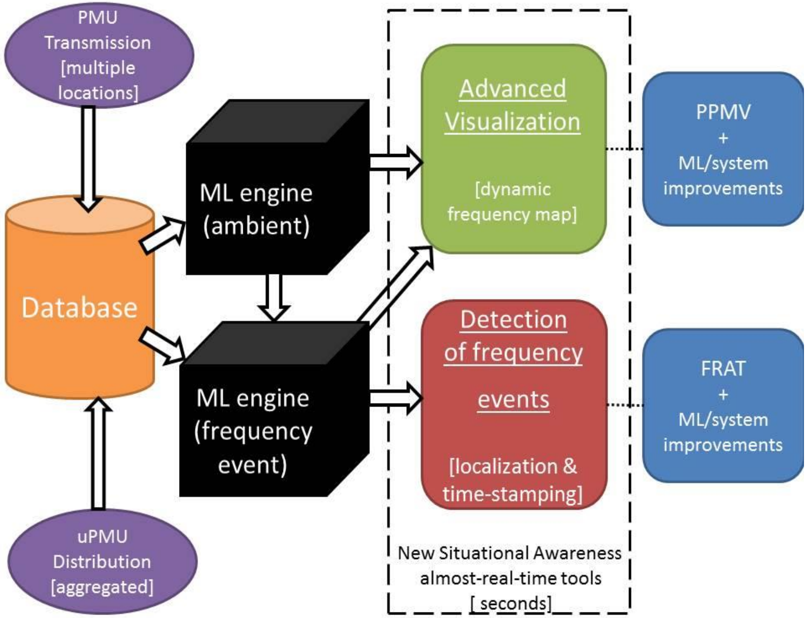
Machine Learning for the Transmission Grid:
Industry-grade Implementation

GMLC 2.0 proposal
in collaboration with
LBNL, PNNL, Columbia U

Goal:

- Develop data aided architecture
- Database of past events
- Combine PMU with SCADA + (aggregated) uPMU

✓ Grid-informed ML Analysis (just discussed) and New Tools (advanced visualization, events detection)
✓ Validation against and developing industry standards
  - Principal Component Analysis
  - Existing software (PPMV, FRAT)
✓ Optimal sizing/sampling of PMUs



- Intro: Overview of Challenges and Approaches
- Technical Intro: Direct and Inverse Stochastic Problem
  –Machine Learning for Grid Operations
- Machine Learning for Distribution Grid
- Machine Learning for Transmission Grid
- Graphical Models & New Physics/Grid Informed ML-tools

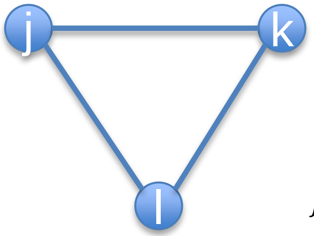
01/18-22/16

cnls.lanl.gov/machinelearning

PHYSICS INFORMED MACHINE LEARNING

Graphical Models for Power Systems (and beyond)

3-bus Power System
v-voltage
s-(apparent) power

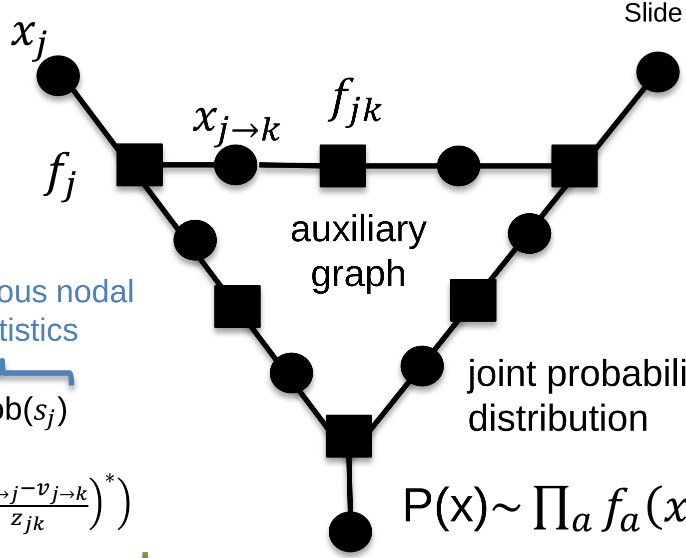


$x_j = (v_j, s_j) \qquad x_{j\to k} = (v_{j\to k}, s_{j\to k})$

a$\in (j, k, l, j \to k, k \to j, k \to l, l \to k, j \to l, l \to j)$

$$f_j(x_j, x_{j\to k}, x_{j\to l}) = I(s_j, s_{j\to k} + s_{j\to l}) * I(v_j, v_{j\to k}, v_{j\to l}) * \overbrace{\text{Prob}(s_j)}^{\text{exogenous nodal statistics}}$$

$$f_{jk}(x_{j\to k}, x_{k\to j}) = \underbrace{I\left(s_{j\to k}, v_{j\to k}\left(\frac{v_{j\to k}-v_{k\to j}}{z_{jk}}\right)^*\right) * I\left(s_{k\to j}, v_{k\to j}\left(\frac{v_{k\to j}-v_{j\to k}}{z_{jk}}\right)^*\right)}_{\text{power flows}}$$



$P(x) \sim \prod_a f_a(x_a)$

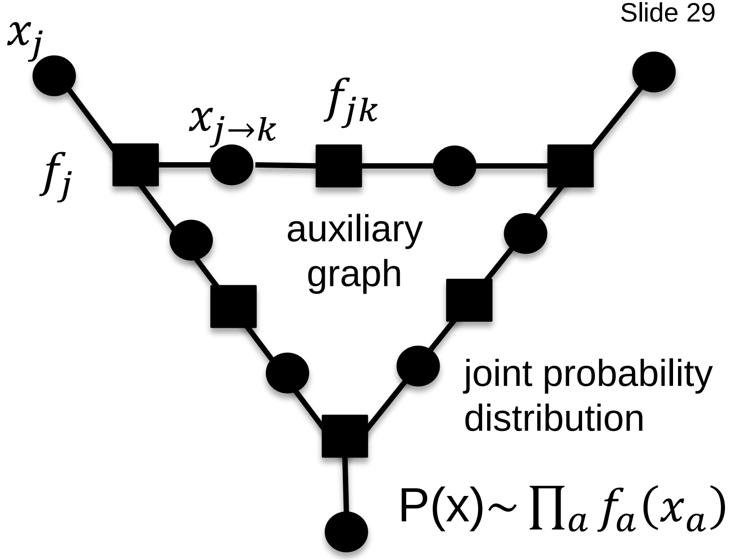
e.g. opens it up for new Machine Learning + solutions

Universal formulations for all statistical objects of Interest:

- Marginal Probability of voltage at a node - $P(v_j) = \sum_{x \backslash v_j} P(x)$
- Most probable load/wind at a node [instanton]
  keeping voltages within a domain - $argmax_{s_j} \sum_{x \backslash (s_j)} P(x)_{v \in Dom_v}$
- Stochastic Optimum Power Flows (CC-, robust-) + dynamic (multi-stage) + planning ++
- Allows to incorporate multiple "*complications*"
  - Any deterministic constraints (limits, inequalities), e.g. expressing feasibility
  - Any mixed (discrete/continuous) variables, e.g. switching

Complexity of Learning: Easy vs Hard

- Direct Problem – Statistical Inference (marginal, partition function, ML)
- Inverse Problem – Learning (graphs & factors) from samples



- New Story (2015) – Don't follow the sufficient statistics path
- Focus on Sample and Computational Complexity of finite GM Learning
- Provably efficient "local" optimization schemes (binary, pair-wise GM)
  - based on ``**conditioning**" to vicinity of a local variable *[Bressler 2015]*
  - based on ``screening" interaction through an accurate choice of the optimization cost *[M. Vuffray, A. Lokhov, S. Misra, MC 2016]*
    - generalizable – applies directly to an arbitrary GM

Summary & Path Forward

- **ML for distribution** – PF-aware spanning tree algorithm to learn structure (forest) and correlations of loads
- **ML for transmission** – two-state on-line learning – ambient + emergency [learning parameters of ODEs, model reduction, waves]
- **Graphical Models** – proper language for variety of stochastic grid problems, e.g. related to learning.
  - Recent progress in GM learning -light, distributed, provably exact schemes – applies naturally to the grid-specific (and other physical network-specific) ML problems.
  - New relaxation ideas based on adaptive Linear Programming – Generalized Belief Propagation schemes – complementary to ``standard" relaxations for OPF & related

LANL Grid Science Team


S. Backhaus


R. Bent


M. Chertkov


C. Coffrin


A. Zlotnik


S. Misra


M. Vuffray


H. Nagarajan


C. Borraz-Sanchez


A. Lokhov


E. Yamangil


D.Deka

Back to main presentation

The Ising Model Learning Problem

Generate M i.i.d. samples

of binary sequences

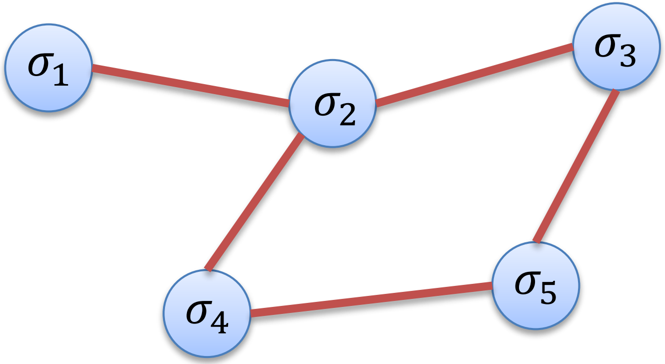
$$\mu(\sigma_1, \dots, \sigma_N) \propto \prod_{(i,j)\in E} \exp\left(J_{ij}\sigma_i\sigma_j\right)$$

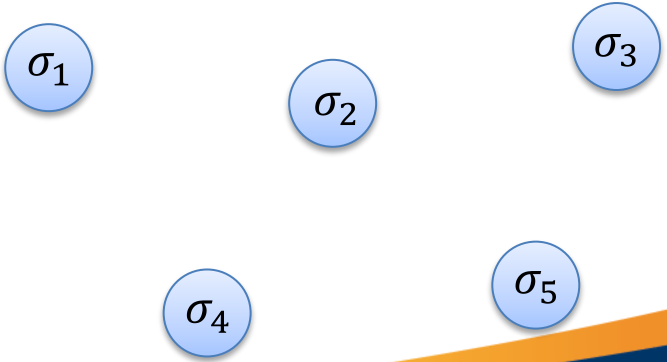
$$\begin{array}{c} \left(\sigma_1^{(1)}, \dots, \sigma_N^{(1)}\right) \\ \vdots \qquad\qquad \vdots \\ \left(\sigma_1^{(M)}, \dots, \sigma_N^{(M)}\right) \end{array}$$

Reconstruct graph and

couplings with high

probability





Back to main presentation

Learning is Easy in Theory & Practice

Number of variables: N

Maximum node degree: d

Number of samples: M

Coupling intensity: $J_{min} \leq |J_{ij}| \leq J_{max}$

Complexity: $\exp\left(\frac{e^{c_1 d J_{max}}}{J_{min}^{c_1}}\right) N^2 \log N$

Samples Required: $\exp\left(\frac{e^{c_1 d J_{max}}}{J_{min}^{c_1}}\right) \log N$

Bresler (2015)
Structure Learning

Complexity: $\frac{e^{8 d J_{max}}}{J_{min}^2} N^3 \log N$

Samples Required: $\frac{e^{8 d J_{max}}}{J_{min}^2} \log N$

Vuffray et al. (2016)
Structure + Parameter Learning

We develop **new model estimators**: (Regularized) Interaction Screening Estimators

They are **consistent estimators** for **all graphical models** (Continuous variables, general interactions, etc…)

Provably optimal on **arbitrary** Ising Models, distributed

Back to main presentation

The Screening Estimator(s)

Number of samples: ∞

$$f_u(\theta) = \left\langle \prod_{j \neq u} \exp(-\theta_{ju}\sigma_j\sigma_u) \right\rangle$$

$$\hat{J}_u = \operatorname*{argmin}_{\theta} f_u(\theta)$$

Number of samples: M

$$f_u^M(\theta) = \frac{1}{M} \sum_{k=1,\dots,M} \prod_{j \neq u} \exp\left(-\theta_{ju}\sigma_j^{(k)}\sigma_j^{(k)}\right)$$

$$\hat{J}_u^M = \operatorname*{argmin}_{\theta} f_u^M(\theta) + \lambda_{N,M} \|\theta\|_1$$

Regularizer reduces # of samples required:
$O(N \ln N) \longrightarrow O(\ln N)$